

# emldetect

February 23, 2011

## Abstract

EPIC maximum likelihood multi-source point spread function fitting.

## 1 Instruments/Modes

Instrument	Mode
EPIC MOS:	IMAGING
EPIC PN:	IMAGING

## 2 Use

pipeline processing	yes
interactive analysis	yes

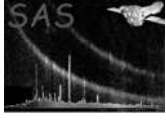
## 3 Description

For an input list of EBOXDETECT source locations, simultaneous maximum likelihood PSF fits to the source count distribution are performed in all energy bands of each EPIC telescope. A description of the main properties of the detection algorithm may be found in Cruddace, Hasinger, Schmitt (1988).

Free fit parameters are: Source location (alpha, delta), source extent (Gaussian sigma or beta model core radius), and source count rates in each energy band of each telescope. The source location and source extent are constrained to the same best-fit value in all energy bands of each EPIC instrument whereas the source count rates are adjusted to their individual best-fit value in each energy band of each EPIC instrument. Derived parameters are: Total source count rate, likelihood of detection (total and in each energy band), likelihood of source extent, and three hardness ratios.

All detection likelihoods are transformed to equivalent likelihoods  $L_2$  (Column DET\_ML of the output source table), corresponding to the case of two free parameters to allow comparison between detection runs with different numbers of free parameters (i.e., when different numbers of input images are used):

$$L_2 = -\ln(1 - P(\frac{\nu}{2}, L')) \quad \text{with} \quad L' = \sum_{i=1}^n L_i$$



where  $P$  is the incomplete Gamma function,  $n$  is the number of energy bands involved,  $\nu$  is the number of degrees of freedom of the fit ( $\nu = 3 + n$  if task parameter **fitextent**=yes and  $\nu = 2 + n$  otherwise), and  $L_i = C_i/2$  with  $C$  as defined by Cash (1979). Note, that  $n=1$  in the case of the individual energy band detection likelihoods listed in source table rows with **ID\_BAND** > 0 and  $n$  is equal to the total number of energy bands in the **ID\_BAND**=0 summary rows. The transformed detection likelihoods obey the simple relationship  $L_2 = -\ln(p)$  where  $p$  is the probability for a random Poissonian fluctuation to have caused the observed source counts. This is in agreement with the detection likelihoods as defined in task **eboxdetect** (column **SIGMA** of **eboxdetect** source lists). Note that for very small numbers of source counts (less than  $\approx 9$  counts, Cash 1979) this relation has to be treated with caution. Therefore it will only give a rough estimate of the number of expected spurious sources.

If the parameter **fitextent** is set to “yes”, the point spread function will be convolved with a source extent model, that can be set to either a Gaussian profile or a  $\beta$ -model profile via the parameter **extentmodel**. In the case of **extentmodel**=beta the surface brightness is calculated as

$$f(x, y) = \left( 1 + \frac{(x - x_0)^2 + (y - y_0)^2}{r_c^2} \right)^{-3/2}$$

The value of the core radius  $r_c$  is written to the column **EXT** of the output source list. In the case of a Gaussian extent model  $\sigma$  is written to column **EXT**.

From version 4.27 also the extent likelihood values (**EXT\_ML**) are corrected for the number of input images with the formalism described above.

If detection over several energy bands is performed, up to three hardness ratios are calculated from the source counts rates in the individual bands. The hardness ratios are defined as follows:

$$HR_i = \frac{B_m - B_n}{B_m + B_n}$$

where  $B$  denotes the count rates in energy bands  $n$  and  $m$ , respectively. The energy bands  $n$  and  $m$  used to calculate the hardness ratios can be specified for each instrument via the parameters **hrpndef**, **hrm1def**, and **hrm2def**. The default band assignments (identical for all instruments) are given in the following table:

$i$	$n$	$m$
1	1	2
2	2	3
3	3	4

All EPIC PN source count rates and fluxes written to the **emldetect** source list are corrected for photons arriving during readout of the PN CCDs and therefore are not detected on the nominal source position (out-of-time events). The correction factor is 1.0626 for PrimeFullWindow mode and 1.0223 for the PrimeFullWindowExtended mode. No correction is applied to data in other observing modes. From **eexpmap** v3.31 on the OOT events correction is applied to the exposure maps in all observing modes. **Emldetect** (v4.28 or higher) reads the keyword **OOTCORR** from the FITS header of the exposure maps. If **OOTCORR** is existing and set to “true”, no further correction is applied by **emldetect**.

The band numbers  $n$ , and  $m$  are assigned to the individual bands by numbering the corresponding input images in the order in which they are given on the command line. It is therefore important that the ordering of the input images is consistent with the contents of **hrdef** to obtain meaningful hardness ratios.



Note that the source extent can only be determined reliably for relatively bright objects. If the likelihood of the source extent falls below a threshold, point source parameters are derived.

For each detected source the output source table contains one row for each energy band of each instrument. In addition, summary rows list combined results per instrument and total. The summary rows over the energy bands for each instruments (`ID_BAND = 0`) contain sums of the entries in the individual energy bands, where appropriate (counts, count rates, fluxes, and detection likelihoods). Spatial parameters (positions and extent values) are identical for all energy bands and repeated in the summary row.

The individual source rows are identified through the column entries `ID_INST` and `ID_BAND` in the output table where `ID_INST` refers to the EPIC instrument (1: PN, 2: MOS1, 3: MOS2, 0: summary row) and `ID_BAND` is the energy band number as defined by the ordering in which the energy bands are given on the command line. An `ID_BAND` value of 0 again refers to the summary information. An `ID_BAND` value of 9 stands for the XID energy band (0.5 - 4.5 keV) which will only be present if the input parameter `withxidband` has been set to true. The upper and lower bounds of each energy band are available in the header keywords `aa_n_ELO` and `aa_n_EHI` where `aa` stands for the EPIC camera (PN, M1, or M2) and `n` stands for the energy band number as given in table column `ID_BAND`. Additional keywords `N_INST`, `aa_BNDS`, and `XID_BND` specify the number of EPIC cameras, number of energy bands for each EPIC camera, and whether XID band information is present in the source table. Note, that the energy bands which constitute the XID band have to be specified for each instrument separately using the parameters `xidpndef`, `xidm1def`, `xidm2def` if the default values (bands 2 and 3, as defined in `ID_BAND` column) are not appropriate.

From v4.42.5 a new method to treat the XID band has been introduced: With the parameter `xidfixed` set, `emldetect` can be run on one XID band image per instrument using an **`emldetect`** output list as input source list. Positions and source extent values will be kept fixed and only fluxes and detection likelihoods are determined. In this case the input images (science images, exposure maps, background images) for the desired band (e.g. 0.5-4.5 keV) have to be prepared beforehand. The parameters `xidpndef`, `xidm1def`, `xidm2def` are used to determine which energy bands from the input source list are used to provide the start values for the fit. Note that for using this method `withxidband` should be set to “no” and parameter `ecf` is used to set the energy conversion factors.

Simultaneous fitting of data from different instruments (i.e., all EPIC pn and MOS data) or different exposures is supported. The PSF fitting may either be performed in single source or in multi-source mode. In multi-source mode sources with overlapping PSFs are fitted simultaneously. Up to six neighbouring sources may be fitted simultaneously. Selection of sources for simultaneous fitting is controlled by a distance parameter (`scut`) and the maximum number of sources to be fit simultaneously (parameter `nmaxfit`). Sources fit simultaneously are identified in the the output table through the `ID_CLUSTER` table column. It is also possible to fit several PSFs (up to three) for each input source position by setting parameter `nmulsou` to the appropriate value.

Since both multi-PSF fitting and extent fitting are CPU intensive, two methods to reduce the CPU requirement of an `emldetect` run using multi-PSF fitting. With the option `withthreshold` the user can limit the application of multi-PSF fitting (as specified by `nmulsou`) to sources exceeding a certain threshold ( set by parameter `threshold` in the input column defined by parameter `threshcolumn`. The column can be `LIKE`, `SCTS`, or `RATE`.

The second method to save CPU time for combined extent and multi-PSF fitting is provided by the option `withtwostage`, which is used in combination with `fitextent="true"` and `nmulsou > 1`. If this parameter is set to “true”, **`emldetect`** will perform the fit for each source in 2 stages: In the first stage one extended source is fitted to the source. Only if the extent is significant, the second stage, where a multi-PSF fit with one extended source and `nmulsou-1` point sources is applied, will be performed. This `withtwostage` option in most cases avoids misidentification of close pairs of point sources as extended sources and significantly reduces CPU time.



If parameter `usecalpsf` is set to true (this is the default) PSF fitting is performed using a tabulated energy and position dependent PSF as provided in the calibration database (accuracy level = medium). Alternatively, an internal, hardcoded PSF as specified in section “Algorithm” is used. The hardcoded PSF consists of a superposition of four Gaussian profiles with an off-axis dependent width. It is axially symmetric and currently has no energy dependence. Right ascension and declination of the optical axis positions used for the calculation of the off-axis angles are stored in keywords with the names `OARAAann` and `OADEAann` where “aa” designates the EPIC camera (PN, M1, or M2) and “nn” is a running number. Note that the  $\beta$ -model source extent function (`extentmodel=beta`) can only be used with the CAL PSF (`usecalpsf=true`).

Two parameters determine the image region, on which a source fit is performed: the parameter `ecut` determines the size of the subimage used for fitting a source. The parameter `scut` determines the radius around each source, in which other input sources are considered for multi-PSF fitting, if parameter `nmulsou` is  $> 1$ . Both `ecut` and `scut` are given as encircled energy fraction of the calibration PSF. The actual radii in pixel units therefore change slightly with energy band and source position. The actual value for the cutout radius for each sources is listed in the column CUTRAD of the output source list.

From version 4.32 the maximum value of the extent fit parameter can be given via the task parameter `maxextent`. The unit is image pixels. Large values of `maxextent` can in some cases lead to spurious detection of extended sources. With the parameter `minextent` the minimum extent, which is still considered as significant, can be specified. If the best fit extent is less than `minextent`, a point source model will be adopted for the source.

For version higher than 5.0 (or equal) the parameter `imagebuffersize` is implemented. The main purpose of this parameter is for the processing of mosaic-pointings, where the sky image will contain large areas with no photon data. The value of `imagebuffersize` then should at least be as large as the size length of the area containing photon data. Its default value is 640, which is appropriate for image bin size of 4".

## 4 References

Cash, W., Parameter estimation in Astronomy through application of the likelihood ratio, ApJ, 228, p. 939 (1979)

Cruddace, R., G., Hasinger, G., Schmitt, J. H., The application of a maximum likelihood analysis to detection of sources in the ROSAT database, in ‘Astronomy from large Databases’, eds. Murtagh, F. and Heck, A., p. 177, (1988).

## 5 Parameters

This section documents the parameters recognized by this task (if any).

Parameter	Mand	Type	Default	Constraints
-----------	------	------	---------	-------------

<b>imagesets</b>	yes	filename list	image.fits	
------------------	-----	------------------	------------	--

Names of input EPIC fits images<sup>1,2</sup> or event lists<sup>3</sup> (if useevents = true; not yet implemented)

<b>expimagesets</b>	no	filename list	expimage.fits	
---------------------	----	------------------	---------------	--

Names of exposure images<sup>1,2</sup>



<b>detmasksets</b>	no	filename list	detmask.fits	
--------------------	----	------------------	--------------	--

Names of detection masks<sup>3</sup>

<b>bkgimagesets</b>	yes	filename list	bkgimage.fits	
---------------------	-----	------------------	---------------	--

Names of background maps<sup>1,2</sup>

<b>boxlistset</b>	yes	filename	eboxlist.fits	
-------------------	-----	----------	---------------	--

Name of eboxdetect source list

<b>mllistset</b>	yes	filename	emllist.fits	
------------------	-----	----------	--------------	--

Name of output emldetect source list

<b>mergedlistset</b>	no	filename	mergedlist.fits	
----------------------	----	----------	-----------------	--

Name of merged source list

<b>mlmin</b>	no	float	10.0	[1.0<param<50.0]
--------------	----	-------	------	------------------

Minimum detection likelihood for inclusion of source in output list

<b>dmlextmin</b>	no	float	10.0	[1.0<param<100.0]
------------------	----	-------	------	-------------------

Required likelihood improvement for source extent

<b>scut</b>	no	float	0.9	[0.4<param<0.99]
-------------	----	-------	-----	------------------

Source selection radius for multi-source fitting (expressed as fraction of the normalized PSF integrated to the desired radius)

<b>ecut</b>	no	float	0.68	[0.4<param<0.99]
-------------	----	-------	------	------------------

Event cut-out radius for multi-source fitting (expressed as fraction of the normalized PSF integrated to the desired cut-out radius)<sup>4</sup>

<b>ecf</b>	no	float	1.0	[0.001<param<1000]
------------	----	-------	-----	--------------------

Energy conversion factors, given in units of  $10^{11}$  counts  $\text{cm}^2 \text{erg}^{-1}$  <sup>1,2</sup>

<b>xidecf</b>	no	float	1.0	[0.001<param<1000]
---------------	----	-------	-----	--------------------

XID band energy conversion factors, given in units of  $10^{11}$  counts  $\text{cm}^2 \text{erg}^{-1}$  <sup>3</sup>

<b>useevents</b>	no	binary	false	
------------------	----	--------	-------	--

Photon mode flag<sup>6</sup>

<b>fitposition</b>	no	binary	true	
--------------------	----	--------	------	--

Fit source positions

<b>fitextent</b>	no	binary	false	
------------------	----	--------	-------	--

Fit source extent

<b>fitcounts</b>	no	binary	true	
------------------	----	--------	------	--

Fit source counts

<b>fitnegative</b>	no	binary	false	
--------------------	----	--------	-------	--

Allow fitted count rates to become negative

<b>determineerrors</b>	no	binary	true	
------------------------	----	--------	------	--

Determine statistical errors



<b>withexpimage</b>	no	binary	true	
---------------------	----	--------	------	--

Use exposure maps

<b>withdetmask</b>	no	binary	false	
--------------------	----	--------	-------	--

Only sky pixels inside the detection mask will be used in the PSF fits if set to true

<b>withoffsets</b>	no	boolean	false	
--------------------	----	---------	-------	--

Controls whether the merged source list will be read

<b>withxidband</b>	no	boolean	false	
--------------------	----	---------	-------	--

Controls whether XID band output will be written

<b>withsourcemap</b>	no	boolean	false	
----------------------	----	---------	-------	--

Controls whether source maps (one per input image) will be written

<b>sourceimagesets</b>	no	filename list	srcmap.fits	
------------------------	----	------------------	-------------	--

Names of source maps (one per input image)

<b>usecalpsf</b>	no	boolean	true	
------------------	----	---------	------	--

True: read PSF from the calibration db; false: use hardcoded PSF<sup>5</sup>

<b>extentmodel</b>	no	string	gaussian	gaussian—beta
--------------------	----	--------	----------	---------------

Model function for source extent

<b>minextent</b>	no	float	1.5	[0.0<param<300.]
------------------	----	-------	-----	------------------

Minimum allowed value for the extent parameter of an extent model in image pixels

<b>maxextent</b>	no	float	20.0	[0.1<param<300.]
------------------	----	-------	------	------------------

Maximum allowed value for the extent parameter of an extent model in image pixels

<b>withhotpixelfilter</b>	no	boolean	false	
---------------------------	----	---------	-------	--

if true, the likelihood contribution of the brightest pixel will be ignored (i.e., detections relying on a single pixel will be disregarded)

<b>nmaxfit</b>	no	integer	1	1,6
----------------	----	---------	---	-----

Max. number of sources to be fit simultaneously

<b>nmulsou</b>	no	integer	1	1,3
----------------	----	---------	---	-----

Allow fit to split up one input source in maximum nmulsou sources

<b>pimin</b>	no	integer	2000	[0<param<20000]
--------------	----	---------	------	-----------------

Lower energy boundaries of exposure images; units: eV

<b>pimax</b>	no	integer	4500	[0<param<20000]
--------------	----	---------	------	-----------------

Higher energy boundaries of exposure images; units: eV

<b>hrpndef</b>	no	integer	1 2 2 3 3 4	0,10
----------------	----	---------	-------------	------

Array of six integer numbers specifying the upper and lower energy band for each of three hardness ratios (PN)

<b>hrm1def</b>	no	integer	1 2 2 3 3 4	0,10
----------------	----	---------	-------------	------

Array of six integer numbers specifying the upper and lower energy band for each of three hardness ratios (MOS1)



<b>hrm2def</b>	no	integer	1 2 2 3 3 4	0,10
Array of six integer numbers specifying the upper and lower energy band for each of three hardness ratios (MOS2)				
<b>xidpndef</b>	no	integer	2 3	0,10
energy bands used as xid band (PN)				
<b>xidm1def</b>	no	integer	2 3	0,10
energy bands used as xid band (MOS1)				
<b>xidm2def</b>	no	integer	2 3	0,10
energy bands used as xid band (MOS 2)				
<b>xidfixed</b>	no	boolean	false	
Run emldetect on XID band image with positions and source extent fixed to input values				
<b>withthreshold</b>	no	boolean	false	
Allow splitting up into multi-PSF fitting only for sources above threshold				
<b>threshold</b>	no	float	20	[1.0<param<50.0]
Value of threshold for multi-PSF fitting				
<b>threshcolumn</b>	no	string	LIKE	LIKE—SCTS—RATE
Column in input list, on which threshold will be applied				
<b>withtwostage</b>	no	boolean	false	
use two stage process for multi PSF (nmulsou $\geq 1$ ) fitting				
<b>imagebuffersize</b>	no	integer	640	100,10000
Parameter that controls memory requirements for raster scan data.				

<sup>1</sup> Space separated list, sorted by instrument and energy band. I.e., energy band 1 to energy band  $n$  of instrument one is followed by energy band one to  $n$  of instrument two.

<sup>2</sup> One per instrument per energy band

<sup>3</sup> One per instrument

<sup>4</sup> A parameter value of 1.0 would thus correspond to an infinite radius.

<sup>6</sup> Photon mode is not yet implemented in the current task version.

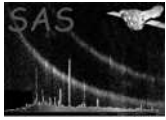
## 6 Errors

This section documents warnings and errors generated by this task (if any). Note that warnings and errors can also be generated in the SAS infrastructure libraries, in which case they would not be documented here. Refer to the index of all errors and warnings available in the HTML version of the SAS documentation.

**MissingParameter** (*error*)

Missing input file name

**WrongInst** (*error*)



Unknown instrument

**FileMismatch** (*error*)

Inconsistent number of input images

**FileMismatch** (*error*)

Inconsistent instruments or bands

**FileMismatch** (*error*)

# detector masks  $\neq$  # instruments

**FileMismatch** (*error*)

Wrong detector mask

**WrongType** (*error*)

Input image has wrong type

**WrongRefPixel** (*error*)

Reference pixel is outside FOV

**wrongParam** (*error*)

ERROR in cut\_radius

**noPSF** (*error*)

point response not valid

**FileMismatch** (*warning*)

Number of PI boundaries not equal number of images

*corrective action:* Use default values for missing boundaries

**FileMismatch** (*warning*)

Number of ECFs not equal number of images

*corrective action:* Use default values

**MissingAttribute** (*warning*)

Keyword is missing

*corrective action:* Keyword is not copied to output

**WrongPointDir** (*warning*)

Pointing direction is outside of image

*corrective action:*

**BufferOverflow** (*warning*)

More than 10000 sources detected

*corrective action:* Remaining sources will be ignored

## 7 Input Files

1. PPS product (from task EVSELECT): EPIC IMAGING mode event lists (one per instrument if program is run in Photon mode)
2. PPS product (from task EVSELECT): FITS images (one per instrument per energy band if program is run in imaging mode)
3. PPS product (from task EBOXDETECT run in map detect mode): EPIC EBOXDETECT source list
4. PPS product (from task EEXPMAP): EPIC exposure images (one per instrument)
5. from task ESPLINEMAP: Spline background images (one per instrument per energy band)





## 8 Output Files

1. PPS product (to be read by task SRCMATCH) : EPIC EMLDETECT source list

## 9 Algorithm

subroutine emldetect

Read in EBOXDETECT source list (map detect) and  
sort by source count rate

Loop over sorted source list (begin with brightest source):

Selection of sources for simultaneous multi-source fitting:

- 1) Search close neighbours within source cut radius of current source. Don't consider sources which have already been processed; mark selected sources as processed.
- 2) Repeat (1) for each close neighbour until the maximum number of sources for simultaneous fitting (max. 8; specified in parameter file) is reached (i.e., the selection of additional sources for the multi-source fitting terminates when either the maximum number of sources (parameter max\\_fit) is reached or when no additional sources fulfill the distance criterium (parameter scut).

Determine data area (2d-mask array) to be used for multi-source fitting:

Loop over mask array: Set to 1 if pixel is (a) within event cut radius of selected source and (b) within the area marked in the detection mask; set to 0 otherwise.

Read in data:

Fill (x, y, count) data records:

Binned mode:

x,y: image pixel coordinates  
count: number of events in pixel

Single photon mode:

x,y: event coordinates  
count:=1

Feed data records, background maps, and exposure maps  
into Maximum Likelihood PSF fitting algorithm:

Use Marquarth-algorithm to minimize likelihood function of  
multi-source PSF fit.

Remove sources from fit which do not significantly improve  
goodness of fit (required improvement in likelihood specified  
by parameter).

Set source extent to 0 if extent does not significantly improve  
goodness of fit (required improvement in likelihood specified  
by parameter).



Add best-fit source models (PSF + source extent) to background maps (i.e., sources which have already been fitted by the program are treated as background for the remaining sources. Note that the sources are processed in the order of decreasing count rate such that all the bright sources will have been modeled into the background map once the weak sources are processed.

```
If likelihood of detection exceeds threshold THEN
  Write source parameters to EMLDETECT source list.
  Add fluxes and hardness ratios.
END IF
```

```
END Loop
```

```
end subroutine emldetect
```

Code segment describing the hardcoded analytical representation of the PSF (used if parameter `usecalpsf` is set to false).

```
! eps:    off-axis angle in arcmin
! sig:    source extent in arcsec
! psf:    surface brightness of point spread function,
!          normalized, such that  $\int_0^\infty 2\pi r dr \text{psf}(r) = 1$  [1/arcsec2]
!          from 0 to infinity = 1 [1/arcsec2]

data norm /0.4169,0.3470,0.1279,0.1082/, &
          sigma /4.1450,8.6218,22.372,66.698/, &
          offsg /0.0324/

! off-axis component of telescope
sigtel=offsg*eps*eps

sum=0.
do i=1,4
  sig2=0.5/(sigma(i)**2+sigtel**2+sig**2)
  rnorm=norm(i)*sig2/pi
  sum=sum+rnorm*exp(-sig2*(x)**2)
end do

psf = sum
```

## 10 Comments

Due to coding error the likelihood values `DET_ML` and `EXT_ML` computed by **emldetect** versions 4.24 and older were overestimated by a factor 2 or more. From version 4.27 the the correct values are computed. Note that this change strongly reduces the number of spurious detections at a certain likelihood threshold.



## 11 Future developments

Photon mode still needs to be implemented.

## References

**Output source table columns**

ML_ID_SRC	EMLDETECT source number
BOX_ID_SRC	corresponding EBOXDETECT input source number
ID_INST	instrument ID; 1: PN, 2: MOS1, 3: MOS2
ID_BAND	energy band number (band number 0: summary band)
ID_CLUSTER	cluster id; sources fit simultaneously have same number
SCTS	source counts
SCTS_ERR	source counts error
X_IMA	source image pixel X coordinate
X_IMA_ERR	error of image pixel X coordinate
Y_IMA	source image pixel Y coordinate
Y_IMA_ERR	error of image pixel Y coordinate
EXT	source extent, gaussian sigma or beta model core radius (image pixel)
EXT_ERR	extent error error
DET_ML	Likelihood of detection
EXT_ML	Likelihood of extent
BG_MAP	background at source location (counts/pixel)
EXP_MAP	exposure at source location (seconds, vignetting corrected)
FLUX	source flux (cgs units)
FLUX_ERR	source flux error error
RATE	source count rate (counts/sec)
RATE_ERR	count rate error
RA	source right ascension (degrees)
DEC	source declination (degrees)
RADEC_ERR	error (arcsec)
LII	source galactic longitude (degrees)
BII	source galactic latitude (degrees)
RAWX	raw X source coordinate
RAWY	raw Y source coordinate
OFFAX	offa axis angle (arcminutes)
CCDNR	chip number
HRn (n=1..4)	hardness ratios 1..4
HRn_ERR (n=1..4)	hardness ratio error
CUTRAD	source cut out radius
MASKFRAC	PSF weighted on-chip fraction
EFF	encircled energy fraction
VIGNETTING	vignetting
ONTIME	Integration time, not vignetting corrected
DIST_NN	distance to nearest neighbour (arcsec)
FLAG	quality flag